

Docket No. AUS920000847US1

**CLAIMS:**

What is claimed is:

5

1. A method in a data processing system for determining an optimal capacity of a server within a set of servers, the method comprising:

10 dynamically collecting resource use and units of work data from the server; and

identifying the optimal capacity for the server using the resource use and unit of work data from the server.

15 2. The method of claim 1 further comprising:

directing connection requests to servers within the set of servers using the optimal capacity for the server.

3. The method of claim 1, wherein the resource use and  
20 unit of work data is collected in response to an event.

4. The method of claim 3, wherein the event is a periodic event.

25 5. The method of claim 1, wherein optimal capacity is identified as follows:

$$E_i = D_{i1}/D_{i2}$$

$$R_i = E_i / (E_1 + E_2 \dots E_n)$$

$$P_i = R_i * 100$$

30 wherein,  $E_i$  is an efficiency of server  $i$ ;  $R_i$  is a relative efficiency of server  $i$ ;  $P_i$  is the optimal

Docket No. AUS920000847US1

capacity of server  $i$ ;  $n$  is a number of servers in the set of servers;  $Di1$  is an average number of units of work handled by server  $i$  since the last time data was sent to the data processing system; and  $Di2$  is an average

5 resource use for server  $i$  since the last time data was sent to the data processing system.

6. The method of claim 1, wherein the set of servers are located in a local area network.

10

7. The method of claim 1, wherein the set of servers are a set of virtual servers located on the data processing system.

15

8. The method of claim 1, wherein the resource use includes at least one of processor use, memory use, and bandwidth use.

20

9. The method of claim 1, wherein the unit of work data includes at least one of a number of packets and a number of connections.

10. The method of claim 1 further comprising:  
sending the optimal capacity to the server.

25

11. The method of claim 10 further comprising:  
sending, by the server, the optimal capacity to a load balancer, wherein the load balancer directs connection requests to the set of servers using the  
30 optimal capacity.

Docket No. AUS920000847US1

12. A method in a data processing system for determining an optimal capacity of the data processing system server, the method comprising:

5 tracking resource use and units of work since a last collection of resource use and units of work data from a server;

sending the resource use and units of work performed data to a server in response to an event; and

10 receiving an identification of an optimal capacity from the server in response to sending the resource use and units of work performed data.

13. The method of claim 12 further comprising:

15 sending the identification received from the server to a load balancer.

14. The method of claim 12, wherein the resource use includes at least one of processor use, memory use, and bandwidth use.

20 15. The method of claim 12, wherein the units of work includes at least one of a number of packets and a number of connections.

25 16. The method of claim 12, wherein the server identifies optimal capacity for a set of servers including the data processing system and further comprising:

30 responsive to an absence of a reception of an identification of the optimal capacity within a selected amount of time, initiating a process to replace the

Docket No. AUS920000847US1

server and perform identifications of optimal capacity for the set of servers.

17. A data processing system comprising:

- 5       a bus system;
- a communications unit connected to the bus, wherein data is sent and received using the communications unit;
- a memory connected to the bus system, wherein a set of instructions are located in the memory; and
- 10       a processor unit connected to the bus system, wherein the processor unit executes the set of instructions to dynamically collect resource use and units of work data from the server; and identify an optimal capacity for the server using the resource use
- 15       and unit of work data from the server.

18. The data processing system of claim 17, wherein the bus system includes a primary bus and a secondary bus.

20   19. The data processing system of claim 17, wherein the processor unit includes a single processor.

20. The data processing system of claim 17, wherein the processor unit includes a plurality of processors.

25

21. The data processing system claim 17, wherein the communications unit is an Ethernet adapter.

22. A data processing system comprising:

- 30       a bus system;
- a communications unit connected to the bus, wherein

Docket No. AUS920000847US1

data is sent and received using the communications unit;

a memory connected to the bus system, wherein a set of instructions are located in the memory; and

a processor unit connected to the bus system,

- 5 wherein the processor unit executes the set of instructions to track resource use and units of work since a last collection of resource use and units of work data from a server; send the resource use and units of work performed data to the server in response to an
- 10 event; and receive an identification of an optimal capacity from the server in response to sending the resource use and units of work performed data.

23. A data processing system for determining an optimal

15 capacity of a server within a set of servers, the data processing system comprising:

collecting means for dynamically collecting resource use and units of work data from the server; and

- identifying means for identifying an optimal
- 20 capacity for the server using the resource use and unit of work data from the server.

24. The data processing system of claim 23 further comprising:

- 25 directing means for directing connection requests to servers within the set of servers using the optimal capacity for the server.

25. The data processing system of claim 23, wherein the
- 30 resource use and unit of work data is collected in response to an event.

Docket No. AUS920000847US1

26. The data processing system of claim 25, wherein the event is a periodic event.

27. The data processing system of claim 23, wherein  
5 optimal capacity is identified as follows:

$$E_i = D_{i1}/D_{i2}$$

$$R_i = E_i/(E_1 + E_2 \dots E_n)$$

$$P_i = R_i * 100$$

wherein,  $E_i$  is an efficiency of server  $i$ ;  $R_i$  is a  
10 relative efficiency of server  $i$ ;  $P_i$  is the optimal  
capacity of server  $i$ ;  $n$  is a number of servers within the  
set of servers;  $D_{i1}$  is an average number of units of work  
handled by server  $i$  since the last time data was sent to  
the data processing system; and  $D_{i2}$  is an average  
15 resource use for server  $i$  since the last time data was  
sent to the data processing system.

28. The data processing system of claim 23, wherein the  
set of servers are located in a local area network.

29. The data processing system of claim 23, wherein the  
set of servers are a set of virtual servers located on  
the data processing system.

25 30. The data processing system of claim 23, wherein the  
resource use includes at least one of processor use,  
memory use, and bandwidth use.

31. The data processing system of claim 23, wherein the  
30 unit of work data includes at least one of a number of  
packets and a number of connections.

Docket No. AUS920000847US1

32. The data processing system of claim 23 further comprising:

5 first sending means for sending the optimal capacity to the server.

33. The data processing system of claim 32, wherein the sending means is a first sending means and further comprising:

10 second sending means for sending, by the server, the optimal capacity to a load balancer, wherein the load balancer directs connection requests to the set of servers using the optimal capacity.

15 34. A data processing system for determining an optimal capacity of the data processing system server, the data processing system comprising:

20 tracking means for tracking resource use and units of work since a last collection of resource use and units of work data from a server;

first sending means for sending the resource use and units of work performed data to the server in response to an event; and

25 receiving means for receiving an identification of an optimal capacity from the server in response to sending the resource use and units of work performed data.

30 35. The data processing system of claim 34, wherein the sending means is a first sending means and further comprising:

Docket No. AUS920000847US1

second sending means for sending the identification received from the server to a load balancer.

36. The data processing system of claim 34, wherein the  
5 resource use includes at least one of processor use, memory use, and bandwidth use.

37. The data processing system of claim 34, wherein the  
10 units of work includes at least one of a number of packets and a number of connections.

38. The data processing system of claim 34, wherein the  
15 server identifies optimal capacity for a set of servers including the data processing system and further comprising:

initiating means, responsive to an absence of a  
reception of an identification of an optimal capacity  
within a selected amount of time, for initiating a  
process to replace the server and perform identifications  
20 of optimal capacity for the set of servers.

39. A computer program product in a computer readable  
medium for determining an optimal capacity of a server  
within a set of servers, the computer program product  
25 comprising:

first instructions for dynamically collecting  
resource use and units of work data from the server; and

second instructions for identifying the optimal  
capacity for the server using the resource use and unit  
30 of work data from the server.



Docket No. AUS920000847US1

40. A computer program product in a computer readable medium for determining an optimal capacity of the data processing system server, the computer program product comprising:

5       first instructions for tracking resource use and units of work since a last collection of resource use and units of work data from a server;

          second instructions for sending the resource use and units of work performed data to the server in response to  
10       an event; and

          third instructions for receiving an identification of an optimal capacity from the server in response to sending the resource use and units of work performed data.